

Research Article

Fuzzy ARTMAP Ensemble Based Decision Making and Application

Min Jin,¹ Zengbing Xu,² Ren Li,² and Dan Wu¹

¹ School of Information Science and Engineering, Hunan University, Changsha, Hunan 410082, China

² Sany Heavy Industry Co., Ltd., Changsha, Hunan 410100, China

Correspondence should be addressed to Min Jin; jinmin@hnu.edu.cn

Received 26 July 2013; Accepted 23 August 2013

Academic Editor: Xudong Zhao

Copyright © 2013 Min Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because the performance of single FAM is affected by the sequence of sample presentation for the offline mode of training, a fuzzy ARTMAP (FAM) ensemble approach based on the improved Bayesian belief method is supposed to improve the classification accuracy. The training samples are input into a committee of FAMs in different sequence, the output from these FAMs is combined, and the final decision is derived by the improved Bayesian belief method. The experiment results show that the proposed FAMs' ensemble can classify the different category reliably and has a better classification performance compared with single FAM.

1. Introduction

Recently, artificial neural networks (ANNs) have been widely used as an intelligent classifier to identify the different categories based on learning pattern from empirical data modeling in complex systems [1]. For example, the BP, RBF, and SVM models have been developed quickly and utilized to classify the different fault classes of the machine equipment [2–6]. However, these traditional neural network methods have limitation on generalization, which can give rise to overfitting models for training samples. To solve the problem, the fuzzy ARTMAP (FAM) neural network is created and applied to the classification field [7–9], which is an incremental and supervised network model and designed in accordance with adaptive resonance theory. Although the FAM is able to overcome the stability-plasticity dilemma [10], in real-world application, the performance of FAM is affected by the sequence of sample presentation for the offline mode of training [11, 12].

For this drawback, some preprocessing procedures, known as the ordering algorithms such as min-max clustering and genetic algorithm [13, 14], have been proposed for FAM. Furthermore, a number of fusion techniques have been proposed for FAM to overcome this problem. Tang and Yan employed the voting algorithm of FAM to diagnose the bearings faults [15], Loo and Rao applied the multiple FAM

based on the probabilistic plurality voting strategy to medical diagnosis and classification problems [16]. Since these voting algorithms do not consider the effect of the number of the sample in each class, an improved Bayesian belief method (BBM) is used to combine multiple FAM classifiers which are offline trained in different sequence of samples in this paper.

In view of the above principles, a novel ensemble FAM classifiers is proposed to improve the classification performance of single FAM. The identification schematic graph is shown in Figure 1. Firstly, through different features extraction methods, some feature parameters are extracted from the raw signals. Secondly, by the modified distance discrimination technique, the optimal feature set is selected from the original feature set. Finally, multiple FAM classifiers ensemble based on the improved BBM is employed to come up with the final classification results. The proposed method is applied to the fault diagnosis of hydraulic pump. The experiment results show the effectiveness of the proposed ensemble FAM classifiers.

2. Fuzzy ARTMAP Ensemble Using the Improved Bayesian Belief Method

2.1. Fuzzy ARTMAP (FAM). FAM consists of two ART modules, namely, ART_a and ART_b modules which are bridged

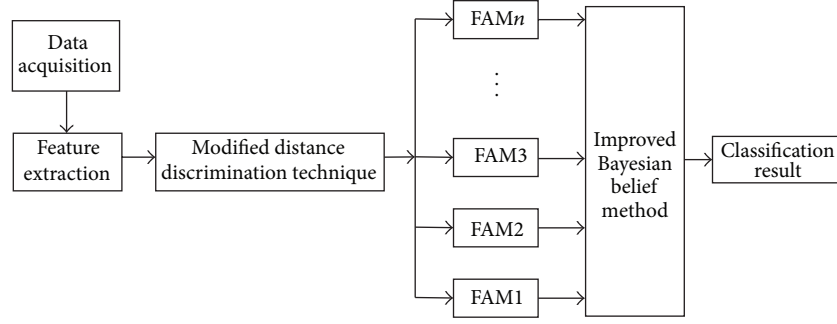


FIGURE 1: Flowchart of classification system.

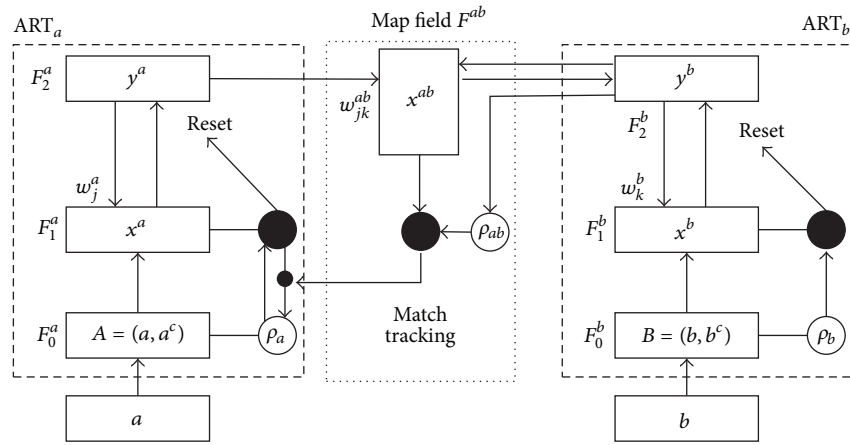


FIGURE 2: The basic structure of fuzzy ARTMAP.

via a map field [10], which is capable to forming associative maps between clusters of input domain in which ART_a module functions as clustering and output domain in which the module ART_b functions as clustering. Each module comprises three layers: normalization layer F_0 , input layer F_1 , and recognition layer F_2 . The structure of FAM is shown in Figure 2. When the output domain is a finite set of class labels, FAM can be utilized as a classifier. The algorithm of FAM can be depicted simply as follows.

The ART_a module receives the input pattern, and the normalization of a M -dimensional input vector \mathbf{a} , is complement-coded to a $2M$ -dimensional vector \mathbf{A}

$$\mathbf{A} = (\mathbf{a}, \mathbf{a}^c) = (a_1, \dots, a_M, 1 - a_1, \dots, 1 - a_M). \quad (1)$$

Then, the dimension of the input vector is kept constant:

$$|\mathbf{A}| = |(\mathbf{a}, \mathbf{a}^c)| = \sum_{i=1}^M a_i + \left(M - \sum_{i=1}^M a_i \right) = M. \quad (2)$$

Afterward, the input sample \mathbf{A} selects the category node stored in the network by the category choice function (CCF):

$$\frac{|A \wedge w_j^a|}{|w_j^a| + \alpha_a} = T_j(A), \quad (3)$$

where \wedge is a min operator, α_a is the choice function of ART_a , and w_j^a is the weight vector of the j th category node.

When a winning category node is selected, a vigilance test (VT), namely, a similarity check against a vigilance parameter ρ_a of the chosen category node, is taken place:

$$\frac{|A \wedge w_j^a|}{|A|} \geq \rho_a, \quad (4)$$

where w_j^a is the winning j th node. When the above category match function (CMF) is satisfied with criterion, the resonance occurs and learning takes place; namely, the weight vector w_j is updated according to the following equation:

$$w_j^{\text{new}} = \beta (A \wedge w_j^{\text{old}}) + (1 - \beta) w_j^{\text{old}}, \quad (5)$$

where $\beta \in [0, 1]$ is the learning rate. Otherwise, a new node is created in F_2^a which codes the input pattern. In the meantime, for the ART_b the same learning algorithm occurs simultaneously using the target pattern.

After the resonance occurs in the ART_a and ART_b , the winning node in F_2^a will send a prediction to ART_b via the map field. The map field vigilance test is used to detect the test. If the test fails, it indicates that the winning node of ART_a predicts an incorrect target class in ART_b ; then a match tracking process initiates. During the match tracking, the value of ρ_a is increased until it is slightly higher than $|A \wedge w_j^a| |A|^{-1}$; then a new search for the other winning node

in ART_a is carried out, and the process continues until the selected F_2^a node can make a correct prediction in ART_b .

2.2. Decision Fusion Using Bayesian Belief Method. The novel Bayesian belief method is supposed in [17]. It is based on the assumption of mutual independency of classifiers and considers the error of each classifier. Assume that in pattern space Z there are M classes and K classifiers. A classifier e_k can be considered as a function:

$$e_k(x) = j, \quad k = 1, 2, \dots, K, \quad j \in \{1, 2, \dots, M, M+1\}. \quad (6)$$

It signifies that the sample x is assigned to class j by the classifier e_k . And its two-dimensional confusion matrix can be represented as

$$CM_k = \begin{bmatrix} n_{11}^k & n_{12}^k & \dots & n_{1M}^k & n_{1(M+1)}^k \\ n_{21}^k & n_{22}^k & \dots & n_{2M}^k & n_{2(M+1)}^k \\ \dots & \dots & \dots & \dots & \dots \\ n_{M1}^k & n_{M2}^k & \dots & n_{MM}^k & n_{M(M+1)}^k \end{bmatrix}, \quad (k = 1, \dots, K),$$

$M+1$ is an unknown label

(7)

which is obtained by executing $e_k(x)$ on the test data set after $e_k(x)$ is trained. Each row i corresponds to class c_i and each column j corresponds to $e_k(x) = j$. The matrix unit n_{ij}^k means the input samples from class c_i while are assigned to class c_j by classifier $e_k(x)$. The number of samples in class c_i is $n_i = \sum_{j=1}^{M+1} n_{ij}^k$, where $i = 1, \dots, M$, and the number of samples labeled j by $e_k(x)$ is $n_j^k = \sum_{i=1}^M n_{ij}^k$, where $j = 1, \dots, M+1$. Considering the difference of the number of samples in each class, on the basis of the confusion matrix a belief measure of classification can be calculated for each classifier by the following belief function [18]:

$$B_k \left(x \in \frac{c_i}{e_k(x)} \right) = P \left(x \in \frac{c_i}{e_k(x)} = j \right) = \frac{n_{ij}^k / \sum_{j=1}^{M+1} n_{ij}^k}{\sum_{i=1}^M (n_{ij}^k / \sum_{j=1}^{M+1} n_{ij}^k)},$$

$i = 1, 2, \dots, M, \quad j = 1, 2, \dots, M+1.$

(8)

When multiple classifiers e_1, e_2, \dots, e_k are developed, their correspondent beliefs B_1, B_2, \dots, B_k are computed based on the performance of base classifiers. Combining the belief measures of all fusion classifiers can result in the final belief measure of the multiple classifier system. In case of equal a priori class probabilities, the combination rule can be depicted as follows:

$$\begin{aligned} B(i) &= B(x \in c_i | e_1(x), e_2(x), \dots, e_k(x), EN) \\ &= P(x \in c_i | e_1(x), e_2(x), \dots, e_k(x), EN) \\ &= \frac{\prod_{k=1}^K B_k(x \in c_i | e_k(x), EN)}{\sum_i \prod_{k=1}^K B_k(x \in c_i | e_k(x), EN)}, \quad i = 1, 2, \dots, M+1. \end{aligned} \quad (9)$$

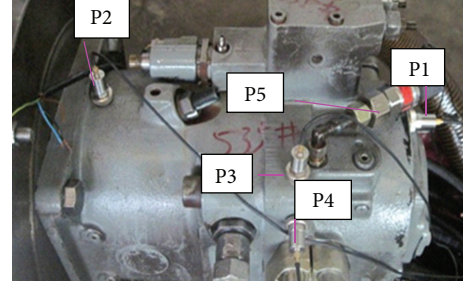


FIGURE 3: The schematic diagram of the experimental setup.

Thus, x is classified into a class j ($j = 1, 2, \dots, M+1$) according to the belief of making the final decision $Bel(j) = \max_{i=1}^{M+1} B(i)$.

3. Case Study

In order to evaluate the effectiveness of the supposed ensemble FAM, the fault identification of hydraulic pump is taken as example. Figure 3 shows the schematic diagram of experiment rig. Four accelerometers are attached to the housing with magnetic bases and mounted at the positions P1, P2, P3, and P4. Pressure sensor is mounted at the position P5. Considering the sensitivity to the fault conditions of hydraulic pump, the vibration signal which is acquired by the accelerometer at the position P2 is utilized to identify the fault categories. And the vibration signals are acquired, respectively, under normal condition and the different fault conditions, such as inner plunger wear, inner race wear, ball wear, swashplate wear, portplate wear, and paraplungers wear.

3.1. Data Preparation. The data set contains 490 samples. These data samples are divided into 245 training and 245 test samples. The detailed descriptions of three data sets are shown in Table 1. In order to identify the different fault categories, a seven-class classification problem need be solved.

3.2. Feature Extraction and Selection

3.2.1. Feature Extraction. Feature parameters are used to characterize the information relevant to the conditions of the hydraulic pump. To acquire more fault-related information, many features in different symptom domains are extracted from the measured signals.

Frequency domain is another description of a signal. In [19], some novel features which can give a much fuller picture of the frequency distribution in each band of frequencies are proposed. Supposed N points of normalized PSD, P_{xx} , of the vibration signal, P_{xx} are divided into L segments, where L is 1 in this study. The four features based on the moment estimates of power can be obtained as follows:

$$\begin{aligned} k_1 &= \frac{1}{N'} \sum_{N'} P_{xx}(n), \\ k_2 &= \frac{1}{N'-1} \sum_{N'} (P_{xx}(n) - k_1)^2, \end{aligned}$$

TABLE 1: Sample subset statistics.

The number of training samples	The number of test samples	Operation condition	Label of classification
45	45	Normal	1
30	30	Inner plunger wear	2
25	25	Inner race wear	3
50	50	Ball wear	4
20	20	Swashplate wear	5
40	40	Portplate wear	6
35	35	Paraplungers wear	7

$$k_3 = \frac{1}{N'k_2^{3/2}} \sum_{N'} (P_{xx}(n) - k_1)^3,$$

$$k_4 = \frac{1}{N'k_2^2} \sum_{N'} (P_{xx}(n) - k_1)^4,$$
(10)

where “ n ” is the number of total data points and N' is the number of sample points in the l th segment.

In order to characterize the spectrum with a higher accuracy, the moment estimates of frequency weighed by power are calculated by the following formulas:

$$k_5 = \frac{1}{K_l} \sum_{N'} f(n) P_{xx}(n),$$

$$k_6 = \sqrt{\frac{\sum_{N'} [(f(n) - k_5)^2 P_{xx}(n)]}{K_l}},$$
(11)

$$k_7 = \frac{1}{K_l k_6^3} \sum_{N'} [(f(n) - k_5)^3 P_{xx}(n)],$$

$$k_8 = \frac{1}{K_l k_6^4} \sum_{N'} [(f(n) - k_5)^4 P_{xx}(n)],$$

where $f(n)$ is the corresponding frequency of $P_{xx}(n)$ and K_l is the total power in the segment. Then, the total number of features extracted for each spectrum is 1×8 .

To depict the fault-related information about the hydraulic pumps quantitatively, the first-order continuous wavelet grey moment (WGM) [20] of vibration signal is extracted. Assuming the wavelet coefficients matrix $[W]_{M \times N}$ which can be displayed by the continuous wavelet transform (CWT) scalogram, M and N are the scales and the time of the scalogram, respectively, the matrix $[W]_{M \times N}$ is divided into m parts along the scale equally, and the first-order wavelet grey moment g_1 of each part can be calculated by the following equation:

$$g_1 = \frac{1}{(M/m) \times N} \sum_{i=1}^{M/m} \sum_j^N w_{ij}^1 \sqrt{(i-1)^2 + (j-1)^2}, \quad (12)$$

where w_{ij} is the element of matrix $[W]_{(M/m) \times N}$. In this paper, the m is set as 8 and the wavelet function is Morlet wavelet.

In addition, due to sensitiveness of these model parameters to the shape of the vibration data, AR model parameters are utilized to characterize the information about the conditions of hydraulic pumps. The AR model is written as follows:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p}, \quad (13)$$

where $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ are the r previous samples, x_t is the predicted sample of the signal, and $\phi_1, \phi_2, \dots, \phi_{t-p}$ is AR model parameters, which can be obtained by the least square method in [21] and expressed by the following formula:

$$\hat{\phi} = (X^T X)^{-1} X^T Y, \quad (14)$$

where

$$X = \begin{bmatrix} x_n & x_{n-1} & \cdots & x_1 \\ x_{n+1} & x_n & \cdots & x_2 \\ & & \ddots & \\ x_{N-1} & x_{N-2} & \cdots & x_{N-n} \end{bmatrix}, \quad (15)$$

$$Y = [x_{n+1}, x_{n+2}, \dots, x_N]^T.$$

In this study, the parameter p is set as 8.

Thus, 24 features constitute the original feature set.

3.2.2. Feature Selection. In order to improve the identification accuracy and reduce the computation burden, some sensitive features providing characteristic information for the classification system need to be selected, and irrelevant or redundant features must be removed. In this study, based on [22], a modified distance discriminant technique is employed to select the optimal features.

Supposing that a feature set of J classes consists of N samples, in the j th class there are N_j samples, where $j = 1, 2, \dots, J$, and $N = \sum_{j=1}^J N_j$. Each sample is represented by M features, and the m th feature of the i th sample is written as f_i^m . Then, the feature selection process can be described as follows.

Step 1. Calculate the standard deviation and the mean of all samples in the m th feature:

$$\sigma_m^2 = \frac{1}{N} \sum_{i=1}^N (f_i^m - \bar{f}^m)^2, \quad \bar{f}^m = \frac{1}{N} \sum_{i=1}^N f_i^m. \quad (16)$$

Step 2. Calculate the standard deviation and the mean of the sample in the j th class in the m th feature, respectively,

$$\sigma_m'^2(j) = \frac{1}{N_j - 1} \sum_{j=1}^{N_j} (f_j^m - \bar{f}_j^m)^2, \quad \bar{f}_j^m = \frac{1}{N_j} \sum_{j=1}^{N_j} f_j^m. \quad (17)$$

Step 3. Calculate the weighted standard deviation of the class center g_j in the m th feature:

$$\sigma_m''^2 = \sum_{j=1}^J \rho_j (g_j^m - g^m)^2 = \mu_1 - \mu_2^2, \quad (18)$$

where $\mu_1 = \sum_{j=1}^J \rho_j (g_j^m)^2$, $\mu_2 = \sum_{j=1}^J \rho_j g_j^m$, $g^m = \sum_{j=1}^J \rho_j g_j^m$, and g^m are the centers of all samples in the m th feature; g_j^m is the center of the samples of the j th class in the m th feature; μ_1 , μ_2 are the weighted means of the squared class center g_j^2 and the class center g_j in the m th feature; ρ_j is the prior probability of the j th class, respectively; and $\sum_{j=1}^J \rho_j = 1$.

Step 4. Calculate the distance discriminant factor of the m th feature:

$$d_b^m - \beta d_w^m = \frac{1}{\sigma_m^2} \left[\sigma_m'^2 - \beta \sum_{j=1}^J \rho_j \sigma_m'^2(j) \right], \quad (19)$$

where d_b^m is the distance of the m th feature between different classes, d_w^m corresponds to the distance of the m th feature within classes, and β is used to control the impact of d_w^m , which is set as 2 in this paper.

Considering the overlapping degree among different classes, a compensation factor is calculated as follows.

Firstly, define and calculate the variance factor of d_w^m in the m th feature as follows:

$$v_w^m = \frac{\max(d_w^m)}{\min(d_w^m)}. \quad (20)$$

Secondly, define and calculate the variance factor of d_b^m in the m th feature as follows:

$$u_b^m = \frac{\max \left\| \bar{f}_i^m - \bar{f}_j^m \right\|}{\min \left\| \bar{f}_i^m - \bar{f}_j^m \right\|}, \quad (21)$$

$$\left\| \bar{f}_i^m - \bar{f}_j^m \right\| = \frac{(\bar{f}_i^m - \bar{f}_j^m)}{\sigma_m^2}, \quad i, j = 1, 2, \dots, J, \quad i \neq j.$$

Then, the compensation factor of the m th feature can be defined and calculated as follows:

$$\eta_m = \frac{1}{v_w^m} + \frac{1}{u_b^m}. \quad (22)$$

Thus, the modified distance discriminant factor can be calculated as follows:

$$\begin{aligned} d_b^m - \beta' d_w^m &= d_b^m - (\beta \eta_m) d_w^m \\ &= \frac{1}{\sigma_m^2} \left[\sigma_m'^2 - \beta \sum_{j=1}^J \left(\frac{1}{v_w^m} + \frac{1}{u_b^m} \right) \rho_j \sigma_m'^2(j) \right], \quad (23) \\ &\quad (m = 1, 2, \dots, M). \end{aligned}$$

Step 5. Rank M features in descending order according to the modified distance discriminant factors $(1/\sigma_m^2)[\sigma_m'^2 - \beta((1/v_w^m) + (1/u_b^m)) \sum_{j=1}^J \rho_j \sigma_m'^2(j)] = \lambda_m$; then normaliz λ_m by $\bar{\lambda}_m = (\lambda_m - \min(\lambda_m))/(\max(\lambda_m) - \min(\lambda_m))$ and get the distance discriminant criteria. Clearly, bigger $\bar{\lambda}_m$ ($m = 1, 2, \dots, M$) signifies that the correspondent feature is better to separate J classes.

Step 6. Set a threshold value γ and select the sensitive features whose distance discriminant factor $\bar{\lambda}_m \geq \gamma$ from the set of M features.

3.3. Diagnosis Analysis. It is well known that the data-ordering of training samples can affect the classification accuracy of single FAM, and that a single output used to represent multiple classes may lead to lower classification accuracy. In order to know how well the proposed FAMs' ensemble work, that is, how significant the generalization ability is improved by utilizing the improved Bayesian belief method to combine the classification results from a committee of single FAM trained with different data-ordering of training samples, the performance of single FAM is also conducted.

In the diagnosis phase performed by the single FAM and FAM ensemble, they are all trained in the fast learning and conservative mode (i.e., setting $\beta = 1$ in (5) and $\alpha_a = 0.001$ in (3)). Besides, in order to ensure the performance of stability-plasticity, the vigilance parameter of FAM is set as $\rho_a = 0.5$, and the ensemble size is set as 5.

In order to improve the classification accuracy and reduce the computation time, in each case some salient features are selected from each feature set by the modified distance discriminant technique, respectively, and then input into the five single FAM in different sequence in the process of training. Figure 4 shows the modified distance discriminant factor $\bar{\lambda}_m$ of all features in the feature sets. From the figure it can be seen that the threshold γ corresponding to the optimal features are different for the case. That is to say, the number of salient features is different.

Figure 5 summarizes the classification results in terms of test accuracy of single FAM and FAMs' ensemble. From the figure, it can be seen that the FAMs' ensemble (0.988) outperforms the single FAMs' in terms of accuracy. And the test accuracy is getting higher when the number of single FAM increases. These indicate that FAMs' ensemble can identify the different fault categories of hydraulic pump well.

3.3.1. Effect of Different Threshold for Feature Selection. As shown in Figure 4, when the threshold value γ is set properly, some redundant and irrelevant features can be removed from the original feature set. To test the effect of the proposed feature selection method based on the modified distance discriminant technique, a series of experiments is carried out against the threshold value γ , in which the parameter of the single FAM is the same as the above, and the size of ensemble FAM is set as 5.

Figure 6 lists the classification accuracy of five individual FAMs' and FAM ensemble against the different thresholds. From the figure, it can be noticed that when $\gamma = 0$ (original feature set), the test accuracy of single FAM and FAM ensemble is 0.824 and 0.845, respectively. The highest test accuracy of single FAM and FAMs' ensemble (0.915 and 0.988) is arrived synchronously when $\gamma = 0.8$, where the optimal feature set is selected. However, when the threshold value continues to increase, the test accuracy of single FAM and FAMs' ensemble tends to decrease. And when threshold $\gamma > 0.9$, the test accuracy of single FAM and FAMs' ensemble

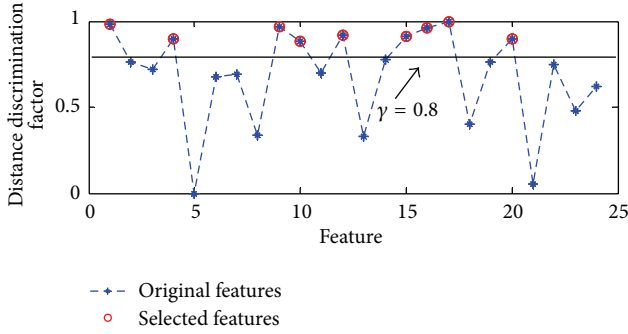


FIGURE 4: Feature selection for three different data sets.

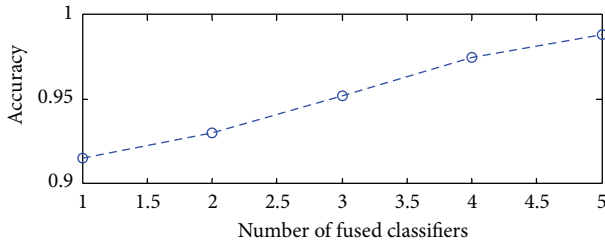


FIGURE 5: Relationship between the test accuracy and the number of single FAM used in FAMs' ensemble.

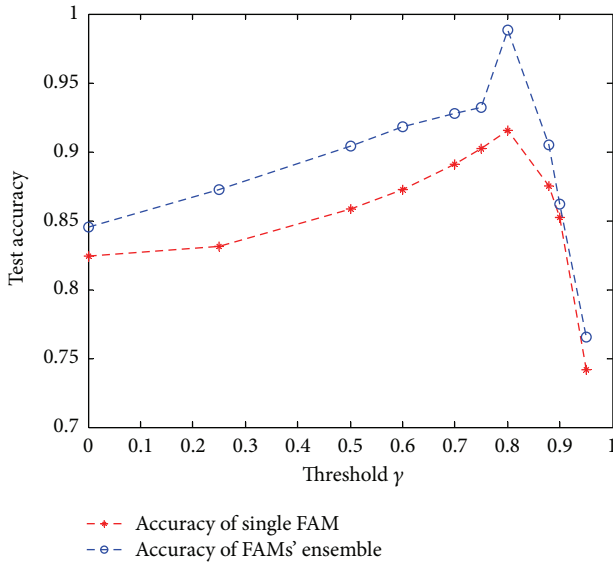


FIGURE 6: Classification accuracy comparison for different threshold.

is lower than that used in all features with threshold $\gamma = 0$. This is mainly because the smaller number of features leads to the overfitting; namely, the drastic reduction of features can lead to a decrease in the test accuracy.

3.3.2. Classification Performance Comparison with Other Classification Methods. In order to test the superiority of the proposed FAMs' ensemble method, the test results produced

TABLE 2: Test accuracy produced by different classification methods.

Method	Test accuracy (%)
Single FAM	91.5
The proposed FAM ensemble	98.8
FAMs' ensemble with voting algorithm and five FAMs	95.2

by FAMs' ensemble and single FAM are compared with those produced by other classification methods. In this experiment, the parameters of FAM ensemble and single FAM are the same as the above.

Table 2 shows the test results of the FAMs' ensemble versus other classification methods. From the table it can be seen that the average test accuracy using single FAMs' is the lowest. However, the test accuracy produced by two FAMs' ensemble methods is higher than that produced by the single classifier, and the test success rate of the proposed FAMs' ensemble is highest and higher than that of FAM ensemble with voting algorithm. These indicate that the proposed FAMs' ensemble has comparatively superior diagnosis performance.

4. Conclusions

The classification performance of FAM is affected by the sequence of training samples. A novel and reliable FAMs' ensemble based on the improved Bayesian belief method is described and proposed to improve the classification performance of FAM in this paper, which combines the output from a committee of FAM fed with different orderings of training samples and derives the combined decision.

And the supposed FAMs' ensemble method is applied to the fault identification of hydraulic pump. The experiment results testify that the proposed FAM ensemble can diagnose the fault categories accurately and reliably and has better diagnosis performance compared with single FAM. These indicate that the proposed FAMs' ensemble has a good promise in the engineering of classification and decision making.

Acknowledgments

This work is supported by the National Scientific and Technological Achievement Transformation Project of China (Grant no. 201255), Electronic Information Industry Development Fund of China (Grant no. 2012407), the National Natural Science Foundation of China (Grant no. 61374172), and the Fundamental Research Funds for the Central Universities, Hunan University, China.

References

- [1] M. M. Polycarpou and A. J. Helmicki, "Automated fault detection and accommodation: a learning systems approach," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 11, pp. 1447–1458, 1995.

- [2] X. Dong, L. Qiu, and Z. Wang, "On line condition monitoring and fault diagnosis for hydraulic pump based on BP algorithm," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 23, no. 3, pp. 322–327, 1997.
- [3] Y. H. Jia, Y. G. Kong, and S. P. Liu, "Application of wavelet neural network to fault diagnosis of hydraulic pumps," in *Proceedings of the 6th International Symposium on Test and Measurement*, pp. 19–22, 2005.
- [4] H. Liu, S. Wang, and P. Ouyang, "Fault diagnosis based on wavelet package and Elman neural network for a hydraulic pump," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 33, no. 1, pp. 67–71, 2007.
- [5] F. Sun and Z. Wei, "Rolling bearing fault diagnosis based on wavelet packet and RBF neural network," in *Proceedings of the 26th Chinese Control Conference (CCC '07)*, pp. 451–455, July 2007.
- [6] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657–665, 2003.
- [7] R. Javadpour and G. M. Knapp, "A fuzzy neural network approach to machine condition monitoring," *Computers and Industrial Engineering*, vol. 45, no. 2, pp. 323–330, 2003.
- [8] S. C. Tan and C. P. Lim, "Application of an adaptive neural network with symbolic rule extraction to fault detection and diagnosis in a power generation plant," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 369–377, 2004.
- [9] X. Zhao, L. Zhang, P. Shi, and H. Karimi, "Novel stability criteria for T-S fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 99, pp. 110–111, 2013.
- [10] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–712, 1992.
- [11] M. Georgiopoulos, H. Fernlund, G. Bebis, and G. L. Heileman, "Order of search in fuzzy ART and fuzzy ARTMAP: effect of the choice parameter," *Neural Networks*, vol. 9, no. 9, pp. 1541–1559, 1996.
- [12] M. Jin, X. Zhou, Z. M. Zhang, and M. M. Tentzeris, "Short-term power load forecasting using grey correlation contest modeling," *Expert Systems with Applications*, vol. 39, no. 1, pp. 773–779, 2012.
- [13] I. Dagher, M. Georgiopoulos, G. L. Heileman, and G. Bebis, "An ordering algorithm for pattern presentation in fuzzy ARTMAP that tends to improve generalization performance," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 768–778, 1999.
- [14] R. Palaniappan and C. Eswaran, "Using genetic algorithm to select the presentation order of training patterns that improves simplified fuzzy ARTMAP classification performance," *Applied Soft Computing Journal*, vol. 9, no. 1, pp. 100–106, 2009.
- [15] Z. Tang and X. Yan, "Voting algorithm of fuzzy ARTMAP and its application to fault diagnosis," in *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '07)*, pp. 535–538, August 2007.
- [16] C. K. Loo and M. V. C. Rao, "Accurate and reliable diagnosis and classification using probabilistic ensemble simplified fuzzy ARTMAP," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1589–1593, 2005.
- [17] L. Lam and C. Y. Suen, "Optimal combinations of pattern classifiers," *Pattern Recognition Letters*, vol. 16, no. 9, pp. 945–954, 1995.
- [18] L. Chen and H. L. Tang, "Improved computation of beliefs based on confusion matrix for combining multiple classifiers," *Electronics Letters*, vol. 40, no. 4, pp. 238–239, 2004.
- [19] M. L. D. Wong, L. B. Jack, and A. K. Nandi, "Modified self-organising map for automated novelty detection applied to vibration signal monitoring," *Mechanical Systems and Signal Processing*, vol. 20, no. 3, pp. 593–610, 2006.
- [20] Z. Yanping, H. Shuhong, H. Jinghong, S. Tao, and L. Wei, "Continuous wavelet grey moment approach for vibration analysis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 20, no. 5, pp. 1202–1220, 2006.
- [21] S. Z. Yang, Y. Wu, and J. P. Xuan, *Time Series Analysis in Engineering Application*, Huazhong University of Science and Technology Press, Wuhan, China, 2007.
- [22] J. Liang, S. Yang, and W. Adam, "Invariant optimal feature selection: a distance discriminant and feature ranking based solution," *Pattern Recognition*, vol. 41, no. 5, pp. 1429–1439, 2008.

